



WORLD HEALTH ORGANIZATION

WHO/HFS/CAS/C/03.65
Distr.: LIMITED
ENGLISH ONLY

**MEETING OF WHO COLLABORATING CENTRES
FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS**

Cologne, Germany
19-25 October 2003

**Title: Lessons Learned throughout the Development of
Electronic Databases for ICD-10-CA and CCI in Two
Languages**

Authors: Diane Caulfeild, B.Sc. P&OT, MBA
Classification Consultant, Canadian Institute for Health Information
Mea Renahan, MBA, CHE
Manager Classification, Canadian Institute for Health Information

WHO Collaborating Center for the Family of International Classifications for
North America

Purpose: For information

Recommendations:

Abstract: (no more than 200 words)

This presentation will highlight some of the lessons learned during the publication cycle of two versions of the CD-ROM in electronic format including the integration of updates, version control, and the additional challenges faced when publishing in more than one language.

In the late 1990's the Canadian Institute for Health Information (CIHI) committed to developing automated processes for the electronic publication of both the ICD-10-CA and CCI (Canadian Classification of Interventions) classification systems. In 2000, a beta version of the CD-ROM was created using an SQL relational database. The first fully functional CD-ROM, Version 2001, was published for use across Canada in March 2001.

In 2002 a data architecture review was undertaken that led to a re-engineering of the relational database. This ensured that all information required for the publication of these classifications was stored within the Oracle 8i database in a way that would simplify the electronic

publication process. A browser based Java web application integrated with a client side XML editor was developed as the maintenance tool for the classification systems. Version 2003 of the CD-ROM with enhancements, addenda and errata corrections was published in March 2003 in both French and English.

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

Lessons Learned throughout the Development of ICD-10-CA and CCI Electronic Databases in Two Languages

The purpose of this paper is to describe the lessons learned throughout the process of creating the International Statistical Classification of Diseases and Related Health Problems – 10th Revision – Canadian Enhancement (ICD-10-CA) and the Canadian Classification of Health Interventions (CCI) in a database format in both French and English. The ICD-10-CA/CCI application supports the two separate classification systems in electronic format.

HISTORY

The development of the electronic product for ICD-10-CA and CCI went through several iterations. The initial prototypes, versions 1.0, 1.1 and 1.2 were MS Word documents on a CD-ROM. The first version of the ICD-10 database was loaded from the Word Perfect documents (ASCII version) provided by the WHO to which the Canadian Institute for Health Information (CIHI) made the enhancements thus creating ICD-10-CA. The CIHI developed the CCI database. The first production of ICD-10-CA and CCI in a database format was Version 2000 using the MS SQL Server database. This was updated as Version 2001 and made available to facilities across Canada for implementation in March 2001 as mandated by the Federal, Provincial and Territorial governments. Version 2003 was developed on the Oracle database and was released across Canada in March 2003. Version 2003 of ICD-10-CA/CCI was produced in French in addition to English for the first time.

With the MS SQL database, much of the information that is integral to the classifications was generated manually rather than generated automatically from the database application. This impacted version control and record management. Text formatting information (e.g. special characters, bullets and braces) was not previously stored in the database and was manually corrected at the time of publication. This prompted the re-engineering of the database from SQL Server 6.5 to Oracle 9i and the application from VB 6.0 to browser-based Java. This was a part of an overall corporate IT strategy at CIHI to provide a more robust, expandable technology base to meet rapidly expanding needs.

The development of the ICD-10-CA and CCI databases was the foundation for the production of an interactive CD-ROM. Our clients across Canada have had the classifications in an electronic book format since April 1, 2001.

VERSION 2001

What did we learn in producing Version 2001?

The Challenges

- The time between the prototype, v. 2000, and the national release of v. 2001 was fewer than twelve months. This short development, production and testing cycle proved to be insufficient to complete a comprehensive Quality Assurance (QA) review.
- There was a steep learning curve for CIHI staff involved in the project. They did not have previous experience in database entry.
- Non-tabular components, such as front and back matter and tables, were not stored in the database. These had to be converted to XML from MS Word or Excel creating an extra step in the process.

- The manual processes associated with the MS SQL database had a high degree of risk for error.
- The SQL Server was not robust enough to efficiently handle demand. This contributed to the corporate decision to move to an Oracle platform.
- The formatting of bullets, braces, notes did not transfer from the database to Folio and had to be entered manually.
- Conversion of the database to Folio infobases and the preparation of the products for electronic, print and PDF distribution were contracted out. Delays occurred throughout the iterative process of working with an external contractor.
- Some of the positive features of the Folio product created some frustrations for our clients when we produced revised infobases (see below).
- Errata, such as typos, jump links not working, missing text were discovered by CIHI and its clients following the release. Most errata were more a nuisance than having a direct impact on the accuracy of the code selection or data submission. This did add to the level of the clients' frustration of having to adopt a whole new system for code identification and data abstraction.

What did we do?

- An on-line coding query service (CQS) was established in June 2001. The primary purpose of the CQS was to address coding queries and through the creation of a searchable database facilitate national data quality. It has also supported our clients as they learned to use electronic coding books, ICD-10 and the Canadian enhancements, CCI and Coding Standards. The added benefit to CIHI is that it provides an ongoing forum for our clients to bring errata to our attention and to make recommendations regarding addenda for future versions.
- Published an errata document on a quarterly basis that was distributed in paper format to every client and also posted on our web site in PDF format.
- Created a Change Order Database to facilitate the tracking and disposition of all identified errata and suggested addenda.
- Released revised infobases, for both ICD-10-CA and CCI in 2002 that fixed all identified errata. The Infobases were available on our website for clients to download and use for submission of their fiscal 2002/03 data.
- Established a National ICD-10-CA/CCI Electronic Products User Group. Representation was sought from every province and territory and embraced both small and large health care facilities. The Terms of Reference of the group include bringing forward not only problems and issues with the electronic products but also proposals for enhancements for future releases. This national forum provides the opportunity for sharing experiences and addressing issues collectively.

What did we learn about our Products?

User Friendly Features of Electronic Code Books

Special features of Folio, such as, the Shadow File and the search engines, and CIHI's custom-built Copy Code Utility tool were advantages in the electronic books. These features facilitated data quality and code look-up.

- The Shadow File is an exact copy or transparency overlay of the Infobase. The Infobase is stored on CD ROM and it is impossible for end users to modify it. This Shadow File feature allows the end-users to add notes to their copy identifying errata and site-specific coding requirements. This feature does have some limitations. A Shadow File cannot be transferred from one Infobase or version to another. In addition, the Shadow File was

unstable and some users experienced problems with corruption of their Shadow Files when multiple users simultaneously worked on it. CIHI issued a patch to address the corruption problem. The issue of transferring the Shadow File from one Infobase to another could not be corrected – it is an inherent characteristic of the Folio product.

- The Copy Code Utility tool was a feature that CIHI had built for copying a code from the Infobase to any other application or document. The intention was to eliminate manual transcription errors by copying and pasting the code into the abstraction application. While this tool has proven to be beneficial it does not support the user who wants to copy multiple codes at one time rather than one code per execution.
- Navigation of the neoplasm index was slow in version 2001. This was overcome in the version 2003 release by building a unique search engine for the neoplasm index.

Revised Infobases - Version 2002

The number of errata and inactive jump-links in v. 2001 of the Folio ICD-10-CA/CCI necessitated the release of revised Infobases in 2002. It is important to note that this was not a re-development of the 2001 databases. The only drawback with this revised 2002 Infobase was that clients were unable to apply their existing Shadow File to the revised Infobases. While numerous notes in the Shadow File, referring to the errata of v. 2001, became redundant; our clients were obliged to rebuild site and user specific notes.

The release of the revised v. 2002 Infobases created a disconnect with the v. 2001 databases at CIHI. The synchronization of the Infobases and the databases was delayed until the development of the v. 2003 databases to allow all CIHI databases to migrate from SQL to Oracle.

VERSION 2003

What new lessons were learned?

The Challenges: Expected and Unexpected

- For version 2003, CIHI had the mandate to produce databases and CD-ROMs in both English and French. The need to synchronize the production and release of the ICD-10-CA/CCI electronic code books and all related products, such as, educational materials and coding standards in both English and French put a further strain on time-lines.
- The transfer of ICD-10-CA/CCI version 2001 databases from SQL to the Oracle platform was three months behind schedule, which delayed the start of database work on v. 2003.
- The time frame to become familiar with the new Oracle based database, learn how to use XML, address the errata, build new codes (over 700 new codes for CCI alone), disable codes, QA and test the applications had now been shortened to nine months.
- There was a requirement to create and load XML documents for the front and back matter, Includes, Excludes, Notes and Indexes in both French and English. This was time consuming.
- The speed of the file server and the capacity of the PC's were inadequate and created further time delays.
- The formatting of bullets, braces, notes and the transfer from the database to Folio continued to be a problem and manual entry was again required.
- With the export of both English and French databases from SQL to XML, time needed to be added into the production schedule to ensure all data were transferred.

What did we do to mitigate the Challenges?

- Conducted a one-day workshop for all staff entering data on XML to demonstrate its functionality with the imported ICD-10-CA/CCI databases.
- The Information Systems department dedicated specialized staff to this project to support staff entering data.
- The French language capacity of the team was enhanced to ensure English and French versions were synchronized.
- To facilitate the data entry into XML fields, a custom JAVA web application was developed for v. 2003 in Oracle Jdeveloper. The JAVA forms allowed the system users to create, modify and disable category codes, maintain code validation edits, conversions and CCI attribute and category components for both the French and the English databases. The JAVA application allows the users to see all the data associated with a code – Includes, Excludes, Notes, graphics and hyperlinks - in a formatted HTML representation similar to the final product. For text entries, the user could choose to edit in the XML in an unformatted text box or download to a user-friendly, customized XML editor. Using the JAVA application for data entry significantly reduced the number of errors in data entry and amount of editing required to generate the Infobases.
- Weekly team meetings were held during the development of v. 2003. All issues were addressed in a timely manner and a status report given by each team member. This proved to be most valuable during the very short, intense time for the development of v. 2003 particularly since all staff were experiencing the same challenges associated with the move to Oracle and XML.
- Clear documentation of accountability and project Gantt charts proved essential to risk management.

Additional Challenges with creating the French Databases

The English 2001 XML documents were translated into French and loaded to the Oracle database. All revisions and addenda to the v. 2003 English database were exported to XML and then translated. After translation, the XML documents were updated back to the Oracle database using an Oracle XSQL utility. A major consideration was the development of the French databases in tandem with the development and enhancements of the English databases. A separate French_Category_Desc Table was created. The French_Desc, long and short titles, were removed and maintained separately with version and date so as not to be a drag on the database. The Language Code added a separate French and English Index. The maintenance to the code links was done twice, since there is not a one-to one relationship between the English and the French databases.

The translation of the English Index entries to the French Index revealed a problem with relationship. It was initially hoped that a one-to-one relationship between the English and French indices would have been possible. The attempt to establish this relationship was to ensure that identical codes were used for linked terms in both languages. It became apparent that it would not be possible to maintain all of these links. The Java application became incredibly slow due to the huge volume of data required each time a drop down list was downloaded from the database during the creation of the Index links between English and French terms. As a result, the Java application was modified to make the linking a totally separate form from the other Index updates.

It also became evident through translation that there were a number of redundant terms in the French Index that needed to be eliminated.

The Results

- The French and English CD-ROMs were released simultaneously as per the contractual stipulation with a bilingual province.
- The release dates for the CD-ROM and related vendor tables provided insufficient lead-time for vendors to make the necessary adjustments to their end-user products. CIHI generates a number of tables in ASCII files for the vendors so they can build their client interfaces. These tables include Code Titles, Conversion Codes and Validation. The short lead times created delays in the delivery of vendor products and created some time delays for some clients in inputting data for the new fiscal year.
- To date, the number of identified errata in the English version is significantly less than v. 2001. It does not appear that a revised Infobase will be required prior to the next release of an updated version.
- A comprehensive QA and testing of the French version is yet to be done. Currently, our bilingual clients are only using the French version as a reference tool. At this time, none of our clients' vendors have developed French versions of their data abstraction software.

GOING FORWARD – THE NEXT VERSION

As we go forward to the next release of the ICD-10-CA/CCI CD-ROM, we have three key areas to focus on. One is a comprehensive review of the French version. We are planning to pilot test this version in our Francophone province prior to its adoption in that province. The second is establishing timelines for the production of the next version of ICD-10-CA/CCI and ensuring these timelines are supportive of other derived CIHI products, such as our grouping methodologies, Canadian Coding Standards and Educational offerings. And, third is meeting the explicit needs of the vendors for building interfaces to other applications and encoders. A production cycle with all interdependencies identified is in the process of development. Once the next release date of all of these products has been established, the production cycle will be adjusted to ensure that all products are available six months prior to the start date of the new fiscal year.

As CIHI goes forward with the development of the next ICD-10-CA/CCI electronic database our plans for improving our processes include - -

- The establishment of two small, dedicated teams, one French the other English, to perform all the production work. These teams will become expert in the demands of building the database thus reducing the probability of errors and increasing quality control.
- Special training in XML and other database components will be made available, as required, to the team members.
- Data entry will be scheduled to commence earlier in the production cycle and be performed at specified time periods (e.g. quarterly), rather than continuously throughout a short, compressed time frame.
- The entire team will be involved in the QA process.
- Weekly status meetings will continue.
- Critical stakeholders within CIHI and the external contractor previously participated only as advisory and ad hoc members of the team. They will be invited to participate as full members of the project team.
- The report writing function of the Change Order database will be enhanced to provide comprehensive reports for the various end-users, such as, the Classification Advisory Committee who review all recommendations for new Canadian codes.
- The feasibility of upgrading the Copy Code Utility tool to allow it to copy multiple codes with one execution will be further explored.

- Additional time will be built into the production cycle to ensure comprehensive QA, including full testing of the functionality of the Infobases prior to release.
- Version 2003 listed in its appendices all new and disabled codes for each classification. These lists were generated from the database. It is proposed that a similar process be developed for the next version to generate a list of the CCI mandatory attributes.
- CIHI is investigating the feasibility of bringing in-house a number of tasks previously contracted out.

CONCLUSION

CIHI has faced several challenges with the development and production of ICD-10-CA/CCI versions 2001 and 2003, in English and French, and has learned from them all. CIHI will continue to work on process controls, and checks and balances to reduce the risk of errata in these products. Our clients have indicated that they believe the ICD-10-CA/CCI electronic coding books and all related supporting products are enhancing the quality and value of Canada's data.